

まこと  
意味的逆引き辞書『真言』  
におけるスコア付け

村脇 有吾 (九大), 栗飯原 俊介 (ACN),  
原田 泰佑 (九大),  
長尾 真 (京大), 田中 久美子 (九大)

# 目次

- 意味的逆引き
- なぜいま逆引きが面白いのか
- 情報検索的ベースライン
- 拡張のための2つの方向性
- 実験
- まとめと今後の課題

# 目次

- 意味的逆引き
- なぜいま逆引きが面白いか
- 情報検索的ベースライン
- 拡張のための2つの方向性
- 実験
- まとめと今後の課題

# 意味的逆引き

# 意味的逆引き

“an instrument that  
indicates directions”

# 意味的逆引き

“an instrument that  
indicates directions”



compass

# 人間用辞書を用いた解法

“an instrument that  
indicates directions”

# 人間用辞書を用いた解法

“an instrument that indicates directions”

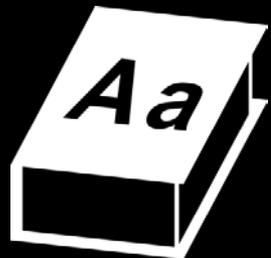


compass ↔

“a device for finding direction with a needle that can move easily and that always points to magnetic north”

# 人間用辞書を用いた解法

compass ↔ “an instrument that indicates directions”



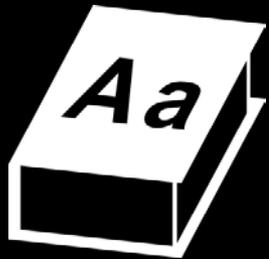
compass ↔ “a device for finding direction with a needle that can move easily and that always points to magnetic north”

# 人間用辞書を用いた解法

compass ↔ クェリ  
“an instrument that indicates directions”



説明文  
“a device for finding direction with a needle that can move easily and that always points to magnetic north”



見出し語  
compass ↔

# 目次

- 意味的逆引き
- なぜいま逆引きが面白いか
- 情報検索的ベースライン
- 拡張のための2つの方向性
- 実験
- まとめと今後の課題

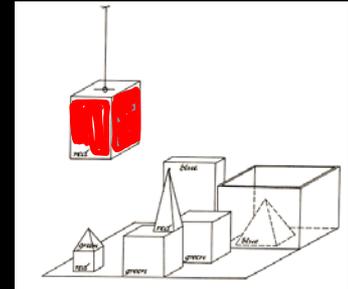
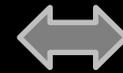
# 逆引きは終わった研究なのか

- 情報検索的手法 [Bilac+, NLP2004]
- WordNet的概念グラフの利用  
[Dutoit+, 2002][Méndez+, 2013][Shaw+, 2013]

# グラウンディング問題とみなせる

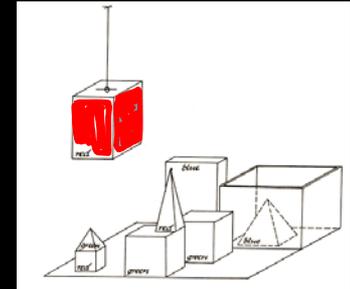
# グラウンディング問題とみなせる

“Pick up a big red block”



# グラウンディング問題とみなせる

“Pick up a big red block”

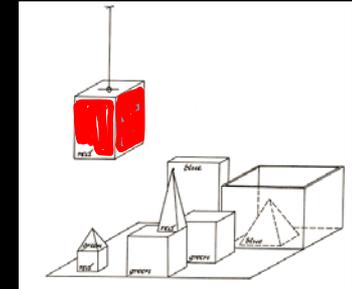


“A green steam train running down the tracks.”



# グラウンディング問題とみなせる

“Pick up a big red block”



“A green steam train running down the tracks.”



“an instrument that indicates directions”



compass

# 分布意味論の研究の進展

- **Words** that occur in the same **contexts** tend to have similar meanings (1950年代)
  - 分布類似度
  - トピックモデル: LSA, LDA, WMF, etc
- 短い**文章**の類似度 [Mihalcea+, 2006][O'Shea+, 2008]
  - Twitterへの応用 [Guo+, 2013]

# 分布意味論の研究の進展

- **Words** that occur in the same **contexts** tend to have similar meanings (1950年代)
  - 分布類似度
  - トピックモデル: LSA, LDA, WMF, etc
- 短い**文章**の類似度 [Mihalcea+, 2006][O'Shea+, 2008]
  - Twitterへの応用 [Guo+, 2013]
- ニューラルな単語表現 [Mikolov+, 2013]
  - King - Man + Woman  $\doteq$  Queen

# 大規模な自動評価が可能

逆引きシステム

システム用辞書



# 大規模な自動評価が可能

評価用辞書

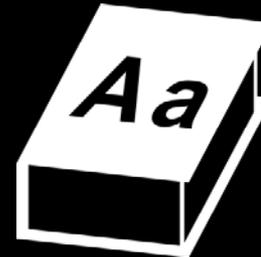


擬似クエリ (説明文)

"A magnetic or electronic device used to determine the cardinal directions."

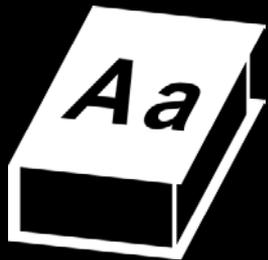
逆引きシステム

システム用辞書



# 大規模な自動評価が可能

評価用辞書



compass ↔

擬似クエリ (説明文)

"A magnetic or electronic device used to determine the cardinal directions."

見出し語候補

1. calculator

2. compass

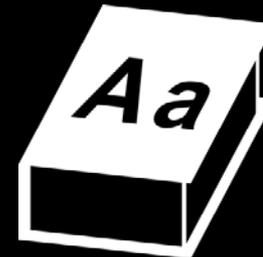
3. avionics

4. firmware

⋮

逆引きシステム

システム用辞書



# 目次

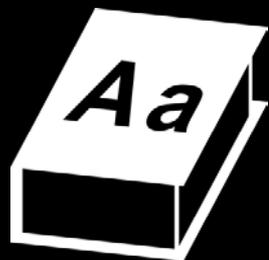
- 意味的逆引き
- なぜいま逆引きが面白いか
- 情報検索的ベースライン
- 拡張のための2つの方向性
- 実験
- まとめと今後の課題

# Bag-of-words表現

compass ↔ クェリ  
"an instrument that indicates directions"



説明文  
"a device for finding direction with a needle that can move easily and that always points to magnetic north"



見出し語  
compass ↔

# Bag-of-words表現

クエリ

“an instrument that indicates directions”



instrument:1,  
indicate:1,  
direction:1



説明文

“a device for finding direction with a needle that can move easily and that always points to magnetic north”



device:1, find:1,  
direction:1,  
needle:1, move:1,  
easily:1, always:1,  
point:1, magnet:1,  
north:1

# Bag-of-words表現

クエリ

“an instrument that indicates directions”



instrument:1,  
indicate:1,  
direction:1



説明文

“a device for finding direction with a needle that can move easily and that always points to magnetic north”



BM25による  
スコアリング

device:1, find:1,  
direction:1,  
needle:1, move:1,  
easily:1, always:1,  
point:1, magnet:1,  
north:1

# 目次

- 意味的逆引き
- なぜいま逆引きが面白いか
- 情報検索的ベースライン
- 拡張のための2つの方向性
- 実験
- まとめと今後の課題

# 拡張1: 単語の並びに着目

compass ↔ クェリ  
"an instrument that indicates directions"



説明文  
"a device for finding direction with a needle that can move easily and that always points to magnetic north"

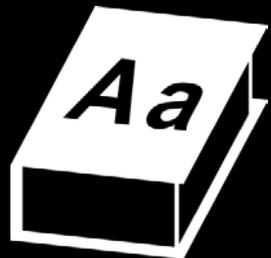


見出し語  
compass ↔

# 拡張1: 単語の並びに着目

compass ↔ クエリ  
"an instrument that indicates directions"

これまで  
クエリ ↔ 説明文  
の対応のみに着目

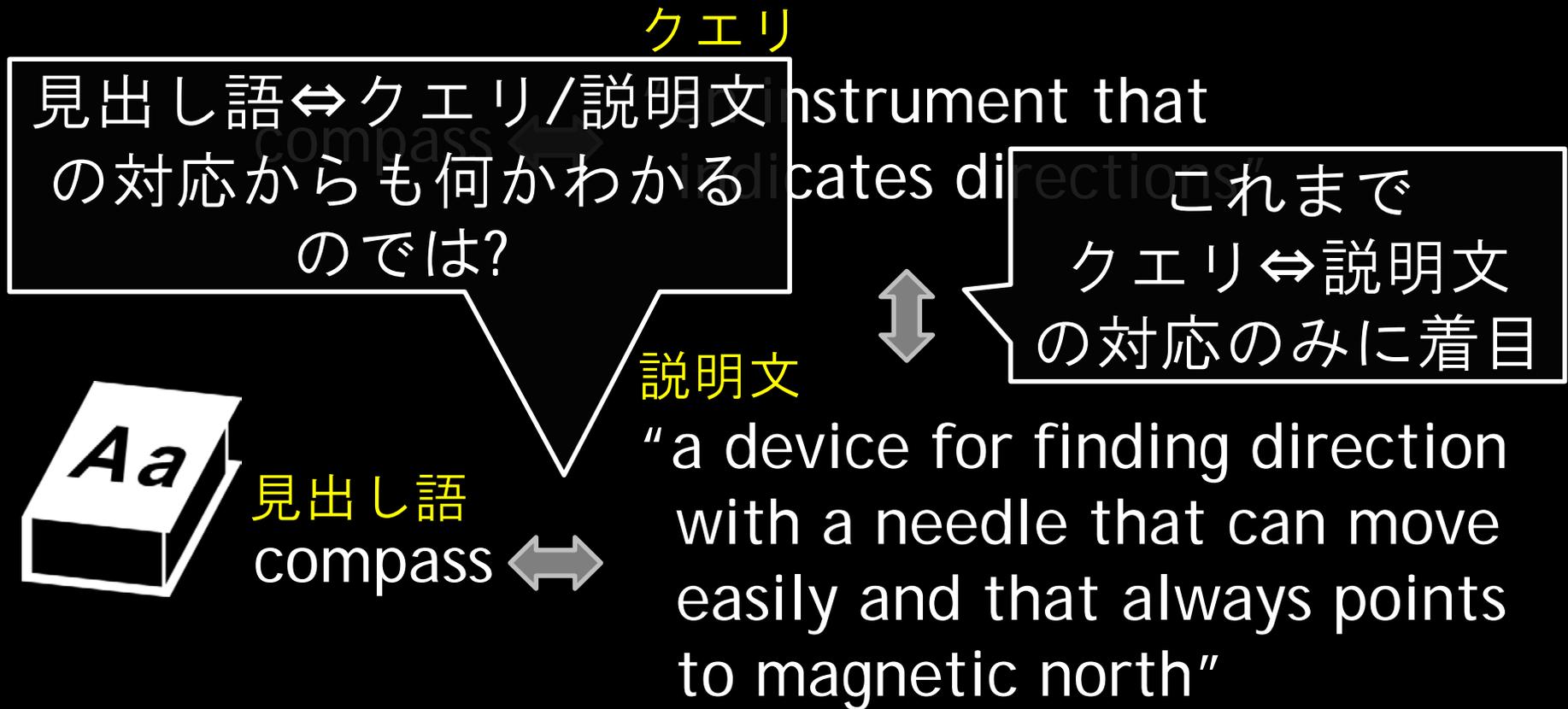


見出し語  
compass ↔

説明文

"a device for finding direction with a needle that can move easily and that always points to magnetic north"

# 拡張1: 単語の並びに着目

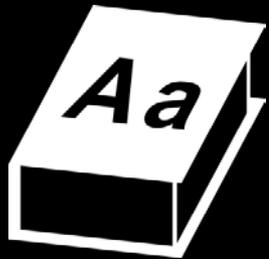


# 拡張1: 単語の並びに着目

compass ↔ クェリ  
"an instrument that indicates directions"



説明文  
"a device for finding direction with a needle that can move easily and that always points to magnetic north"



見出し語  
compass ↔

# 拡張1: 単語の並びに着目

compass ↔ クェリ  
“an instrument that indicates directions”



説明文  
“a device for finding direction with a needle that can move easily and that always points to magnetic north”



見出し語  
compass ↔

# 拡張1: 単語の並びに着目

compass ↔ クェリ  
“an instrument that indicates directions”



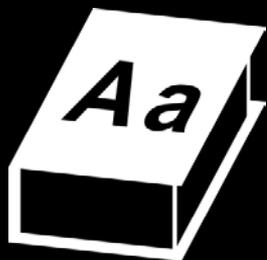
説明文  
“a device for finding direction with a needle that can move easily and that always points to magnetic north”



見出し語  
compass ↔  
品詞: *noun*

# 拡張1: 単語の並びに着目

- 辞書の品詞ラベルを教師として品詞分類器を学習
- 分類結果をsoftmax関数にかけてクエリ・説明文の意味表現に



見出し語

compass

品詞: *noun*

説明文

"a device for finding direction with a needle that can move easily and that always points to magnetic north"

## 拡張2: 単語の意味表現

	instrument	device	direction	south	...
BOW (	1,	0,	1,	0,	... )
BOW (	0,	1,	1,	0,	... )

# 拡張2: 単語の意味表現

>30K

	instrument	device	direction	south	...
BOW (	1,	0,	1,	0,	... )
BOW (	0,	1,	1,	0,	... )

# 拡張2: 単語の意味表現

>30K

	instrument	device	direction	south	...
BOW (	1,	0,	1,	0,	... )
BOW (	0,	1,	1,	0,	... )

単語の意味表現

500

0.15	0.12	-0.64
0.36	0.45	0.05
0.07	-0.02	-0.12
-0.01	0.01	0.09
⋮	⋮	⋮

# 拡張2: 単語の意味表現

>30K

	instrument	device	direction	south	...	
BOW (	1,	0,	1,	0,	... )	クエリ/
BOW (	0,	1,	1,	0,	... )	説明文の 意味表現

単語の意味表現

500

0.15	0.12	-0.64	0.23
0.36	0.45	0.05	0.90
0.07	-0.02	-0.12	-0.56
-0.01	0.01	0.09	0.02
⋮	⋮	⋮	⋮



# 拡張2: 単語の意味表現

文脈の与え方と意味 (直感的な理解)



隣接单語

POS induction



単語ウィンドウ

e.g., word2vec



文書内共起

e.g., LDA

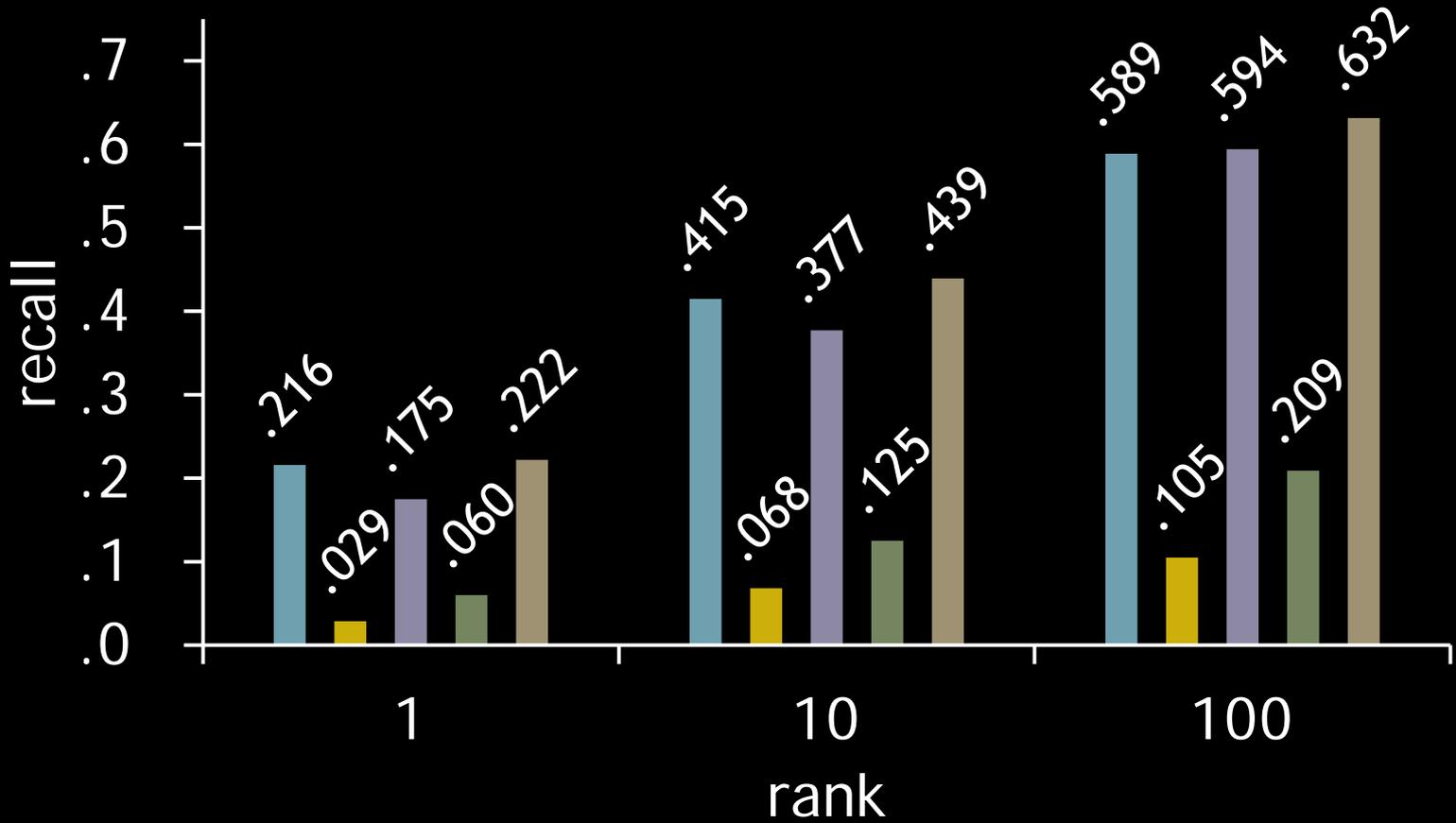
# 目次

- 意味的逆引き
- なぜいま逆引きが面白いのか
- 情報検索的ベースライン
- 拡張のための2つの方向性
- 実験
- まとめと今後の課題

# 実験

- 辞書ペアを用いた自動評価
- システム用辞書: 広辞苑
- 評価用辞書: EDRの日本語単語辞書
- 予稿集ではその他3ペアでも評価
  
- word2vecはWikipediaで訓練

# Recall@N



■ BM25 ■ CAT ■ W2V ■ LDA ■ BM25+CAT+W2V

# 結果

- 単語マッチング (BM25) が強い
- word2vecはデータによってBM25に勝ったり負けたり
- LDAは精度が低い
  - 推定された分布がpeakyすぎる
  - 技術的な問題か、単語内共起の本質的問題か
- 品詞推定は予想通り単独での性能は低い
- スコアの線形和をとると精度があがった

# まとめと今後の課題

- 逆引きは意味研究の良いタスク
- 単語マッチングのベースラインが強い
- 単語の並び & 単語の意味表現に補完的効果
- 今後の課題:
  - 単語の並びと意味表現の統合的モデル

